(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/36977 A2

(51) International Patent Classification[7]: G01N 33/574

(21) International Application Number: PCT/US00/31492

(22) International Filing Date:
16 November 2000 (16.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/165,673	16 November 1999 (16.11.1999)	US
60/172,170	17 December 1999 (17.12.1999)	US
60/178,860	27 January 2000 (27.01.2000)	US
60/201,721	3 May 2000 (03.05.2000)	US
Not furnished	10 November 2000 (10.11.2000)	US

(71) Applicant: MATRITECH, INC. [US/US]; 300 Nevada Street, Newton, MA 02460 (US).

(72) Inventor: WATKINS, Brynmor, A.; 99 Summer Street, Waltham, MA 02452 (US).

(74) Agent: BRODOWSKI, Michael, H.; Testa, Hurwitz & Thibeault, LLP, High Street Tower, 125 High Street, Boston, MA 02110 (US).

(81) Designated States *(national)*: AU, CA, JP.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/36977 A2

(54) Title: METHODS AND COMPOSITIONS FOR IDENTIFYING DISEASE MARKERS

(57) Abstract: The invention provides mass spectroscopic-ased methods and compositions for identifying disease markers, for example, cancer markers, in a mammal. In particular, the mass-spectroscopic-based methods and compositions facilitate the rapid identification of cancer markers from tissue or body fluid samples. The cancer markers, once identified, can then be used as targers in methods for detecting or treating cancer in the mammal.

METHODS AND COMPOSITIONS FOR IDENTIFYING DISEASE MARKERS

Reference to Related Applications

This application claims priority to utility patent application identified by Attorney Docket No. MTP-026, entitled "Methods and Compositions for Identifying Disease Markers," filed on November 10, 2000, and the benefit of U.S. Serial No. 60/165,673, filed November 16, 1999; U.S. Serial No. 60/172,170, filed December 17, 1999; U.S. Serial No. 60/178,860, filed January 27, 2000; and U.S. Serial No. 60/201,721, filed May 3, 2000, the disclosures of which are incorporated by reference herein.

Field of the Invention

The present invention relates generally to methods and compositions for identifying disease markers, for example, cancer markers, in a mammal. More specifically, the present invention relates to mass spectrometry-based methods and compositions for identifying cancer markers in a body fluid.

Background of the Invention

There is an ongoing need to identify new biological markers useful in the detection and/or treatment of various mammalian disorders, for example, cancer. Although a variety of markers have been identified for certain diseases, there is still the need to identify markers for a disease for which no markers presently are available, as well as new markers that are more sensitive and reliable than currently existing markers.

Biochemical markers can be identified by analyzing tissue or body samples from a mammal with the disease of interest and then comparing the results of the analysis with those obtained from a mammal without the disease. One successful approach using two-dimensional gel electrophoresis has led to the identification of a variety of marker proteins that are present at a higher concentration in tissue or body fluid samples of a diseased mammal relative to a normal mammal. See, for example, Partin *et al.* (1993) CANCER RES. 53:744-746 which describes the identification of prostate cancer markers and Getzenberg *et al.* (1996) CANCER RES. 56:1690-1694, which describes the identification of bladder cancer markers.

U.S. Patent No. 5,858,683 discloses a method for identifying cervical cancer in an individual. In the method, protein extracts from samples of normal cervical tissue were

fractionated by two-dimensional gel electrophoresis. Similarly, a second protein extract from samples of cervical cancer biopsy tissue were also fractionated by two-dimensional gel electrophoresis. The resulting gels were compared and spots corresponding to proteins present in higher concentrations in the cancer sample versus the normal sample were identified. Proteins were eluted from the spots of interest on the two-dimensional gel and subjected to conventional protein microsequencing to identify the protein within the spot of interest. This approach has lead to the identification of at least two cervical cancer markers, referred to in the art as TDP-43 and IEF-SSP-9502. Although this approach can be successful, there is still the need to develop a protocol for the more rapid identification of cancer markers and for identifying markers which otherwise may not be detectable using the gel electrophoresis approach.

More recently, an alternative non-electrophoretic-based method (i.e., does not require an electrophoresis step) for identifying cancer markers has been reported in Chang *et al.* (1999) RAPID COMMUN. MASS SPECTRUM. 13, 1808-1812. Lysates from cultured cells (either normal breast cells or malignant breast cells) were fractionated by non-porous reverse-phase high performance liquid chromatography to give protein separation profiles. The more abundant proteins specifically present in the malignant cell lysates were harvested and analyzed by matrix-assisted laser desorption/ionization (MALDI) to determine the masses of the abundant proteins. In addition, a sample of each protein was trypsinized and the tryptic fragments subjected to MALDI to give masses of the fragments which were then compared to protein databases to identify the abundant proteins in the cancer cell based samples. Practice of this method permitted the identification of various proteins, for example, the phosphoprotein p53, the proto-oncogene tyrosine kinase SRC (C-SRC), the c-myc promoter protein and the breast epithelial antigen BA46, all of which were more abundant in the breast cancer lysates. The usefulness of this type of approach for analyzing samples more complex than cell lysates still needs to be evaluated.

There is, therefore, still a need in the art to develop new methods and compositions that can be used to rapidly identify disease markers present in actual tissue or body fluid samples. It is contemplated that such a new method can supplement the already existing methods for identifying disease markers so that additional disease markers can be identified.

Summary of the Invention

The invention provides methods and compositions for the rapid detection and characterization of disease markers, for example, cancer markers, in a mammal, for example, a human. Once identified the markers can be used as targets in assays for detecting the disease, as targets for treatment of the disease or both.

In one aspect, the invention provides a method for identifying a marker molecule indicative of a disease in a mammal. The method comprises the steps of: (a) removing at least one abundant protein from a sample harvested from a mammal with the disease; (b) fractionating the resulting sample depleted of abundant protein to produce a plurality of fractions, each fraction comprising a plurality of molecules; (c) then, separating by mass the molecules disposed within a pre-selected fraction; (d) repeating steps (a) through (c) with a sample harvested from a mammal without the disease; and (e) comparing the molecules separated from the sample from the mammal with the disease with those separated from the sample from the mammal without the disease. As a result, it is possible to rapidly identify one or more marker molecules present at a higher concentration in the sample from the mammal with the disease relative to the sample from the mammal without the disease, wherein the presence of marker molecule is indicative of the disease.

In a preferred embodiment, the sample can be either a tissue or body fluid sample. Preferred body fluids include, for example, blood, serum, plasma, sweat, tears, urine, peritoneal fluid, lymph, vaginal secretion, semen, spinal fluid, ascitic fluid, saliva, sputum, or breast exudate. Serum, however, currently is most preferred.

It has been discovered that by removing one or more abundant proteins from the sample, it is easier to evaluate less abundant proteins as possible disease markers. As used herein, an abundant protein comprises greater than about 5% (w/w), more preferably greater than about 20% (w/w) of total protein in the sample. When the sample is serum, the abundant protein typically is immunoglobulin or albumin. In a preferred embodiment, both immunoglobulin and albumin are removed from the serum to produce an immunoglobulin and albumin depleted serum suitable for further processing.

After depleting the samples of at least one abundant protein, the resulting sample then is fractionated to give a plurality of fractions, with each fraction comprising a plurality of molecules. In a preferred embodiment, the initial fractionation is by a non-electrophoretic

method, for example, by chromatography, more specifically by affinity chromatography. In a more preferred embodiment, the affinity chromatography is ion exchange chromatography, for example, anion exchange chromatography. During ion exchange chromatography, the sample of interest is combined with an appropriate matrix, for example, an anionic or cationic exchange matrix, and molecules are allowed to bind to the matrix. After washing to remove unbound material, the bound molecules then are eluted selectively into different elution buffers, each buffer preferentially eluting a different population of molecules. In ion exchange chromatography, for example, the elution buffers can contain different salt concentrations to permit preferential elution of different types of molecules. It is contemplated that by choosing appropriate buffers it is possible to generate a plurality of fractions, each comprising a plurality of molecules. Alternatively, the affinity chromatography may be performed using a solid support having carbohydrate binding moieties, for example, lectin, disposed thereon. As a result, it is possible to separate carbohydrate containing molecules, for example, glycosylated molecules from non-glycosylated molecules.

One or more of the resulting fractions can then be analyzed by mass-spectroscopy to give the mass of the molecules disposed within a particular fraction. For example, each fraction can be analyzed by matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectroscopy or, more preferably, by surface enhanced laser desorption/ionization-time of flight (SELDI-TOF) mass spectroscopy. During this protocol, the molecules are separated by mass. As a result, it is possible to produce a profile of masses within the sample. By comparing the molecules present at a higher concentration in a sample from a mammal with the disease relative to those present in a sample from a mammal without the disease, it is possible to identify the molecules that are found at elevated levels in the diseased mammal.

If necessary, it is possible to further identify the marker molecules. Further analysis may comprise isolating the molecule and, for example, if the molecule is a protein, then the protein can be further identified by conventional tryptic mapping and/or amino acid sequencing methodologies.

It is contemplated that the method of the invention is particularly effective at identifying markers when the disease is cancer. Accordingly, it is contemplated that the method can be used to identify markers for breast cancer, lung cancer, prostate cancer, bladder cancer, cervical

cancer, ovarian cancer, colon cancer or colorectal cancer. The Examples hereinbelow disclose the identification of breast cancer markers.

In another aspect, marker proteins, once identified, can be used in an assay for diagnosing the disease in a mammal. In a preferred embodiment, the method comprises the steps of: (a) contacting a sample from the mammal with a binding moiety that binds specifically to a disease-associated protein to produce a binding moiety-disease-associated protein complex, wherein the binding moiety binds specifically to a marker protein identified by the method of the invention; and (b) detecting the presence of the complex, which if present is indicative of the presence of disease in the mammal.

In a preferred embodiment, the binding moiety is an antibody, for example, a monoclonal antibody, a polyclonal antibody, or fragment thereof, for example, an Fv, Fab, Fab', $(Fab')_2$ or a biosynthetic antibody binding site, for example, an sFv. The binding moiety preferably is labeled with a detectable moiety, for example, a radioactive label, a hapten label, a fluorescent label, or an enzymatic label.

The presence or amount of the marker protein can thus be indicative of the presence of the disease in the individual. For example, the amount of marker protein in the sample may be compared against a threshold value previously calibrated to indicate the presence or absence of the disease, wherein the amount of the complex in the sample relative to the threshold value can be indicative of the presence or absence of the disease in the individual. Such methods can be performed either on tissue, for example, breast tissue, or a body fluid, for example, serum.

These and other numerous additional aspects and advantages of the invention will become apparent upon consideration of the following figures, detailed description, and claims which follow.

Description of the Drawings

The invention can be more completely understood with reference to the following drawings, in which:

Figures 1 A-C are spectra resulting from the characterization via mass spectrometry of 28 kD protein eluted from a polyacrylamide gel and applied to a nickel SELDI chip. Figure 1A is a spectrum of the heaviest 28 kD protein isolated from the gel, Figure 1B is a spectrum of the

median 28 kD protein isolated from the gel, and Figure 1C is a spectrum of the lightest 28 kD protein isolated from the gel.

Detailed Description of the Invention.

The present invention provides methods and compositions for the identification of disease markers useful as targets either in assays for the detection of the disease or in treatment of the disease. If the marker is, for example, a protein, it is contemplated that the presence of the disease in an individual can be detected using the marker protein and/or binding moieties (*e.g.* antibodies) that bind to the marker protein or to nucleic acid probes which hybridize to nucleic acid sequences encoding the marker protein. Furthermore, it is contemplated that the skilled artisan may produce novel therapeutics for treating the disease which include, for example: antibodies that can be administered to an individual and bind to and reduce or eliminate the biological activity of the target protein *in vivo*; nucleic acid or peptidyl nucleic acid sequences that hybridize with genes or gene transcripts encoding the target proteins thereby to reduce expression of the target proteins *in vivo;* or small molecules, for example, organic molecules which interact with the target proteins or other cellular moieties, for example, receptors for the target proteins, thereby to reduce or eliminate biological activity of the target proteins.

Set forth below are methods for identifying disease markers and methods for detecting the disease by using the marker proteins as targets.

1. Methods for Identifying Disease Markers.

In general, the disease markers are identified by comparing the composition of a sample of tissue or body fluid of a mammal diagnosed with the disease against the composition of a sample similarly treated from an individual without the disease. Accordingly, the resulting markers can be used in assays to detect the presence or absence of a disease in a mammal. Furthermore, it is contemplated that the same method may be employed to identify markers that are present at higher concentrations in one disease state relative to another disease state, for example, an aggressive cancer versus a quiescent cancer.

As used herein, the term "marker" is understood to mean any biological marker, for example, a protein or nucleic acid, which is detectable at a higher level in a tissue or body fluid sample of an individual diagnosed with or diagnosable as having a disease relative to a tissue or body fluid sample of an individual free of the disease and includes species and allelic variants

thereof and fragments thereof. The terms "marker" and "target" are used interchangeably herein. It is not necessary that the marker be unique to a disease state; rather the marker should have a signal to noise ratio high enough to discriminate between samples originating from a diseased individual and samples originating from an individual without the disease.

In one embodiment, the method of the invention comprises the steps of: (a) removing at least one abundant protein from a sample harvested from a mammal with the disease; (b) fractionating the resulting sample depleted of abundant protein to produce a plurality of fractions, each fraction comprising a plurality of molecules; (c) then, separating by mass the molecules disposed within a pre-selected fraction; (d) repeating steps (a) through (c) with a sample harvested from a mammal without the disease; and (e) comparing the molecules separated from the sample from the mammal with the disease with those separated from the sample from the mammal without the disease. As a result, it is possible to rapidly identify one or more marker molecules present at a higher concentration in the sample from the mammal with the disease relative to the sample from the mammal without the disease. The resulting markers, once identified, can be used in an assay to detect the presence or status of a disease, or as a target for therapy.

It is contemplated that the method can be used to identify markers in tissue or body fluid samples. The method, however, is particularly useful in the identification of disease markers in a body fluid, for example, in blood, serum, plasma, sweat, tears, urine, peritoneal fluid, lymph, vaginal secretion, semen, spinal fluid, ascitic fluid, saliva, sputum, or breast exudate. Serum, however, is most preferred.

By removing one or more abundant proteins from the sample, it is easier to evaluate less abundant proteins as possible disease markers. As used herein, an abundant protein comprises greater than about 5% (w/w), more preferably greater than about 20% (w/w) of total protein in the sample. When the sample is serum, the abundant protein typically is immunoglobulin or albumin. It has been reported that in serum, albumin constitutes about 57-71% of total serum protein and that immunoglobulin constitutes 8-26% of total serum protein (Lollo *et al.* (1999) ELECTROPHORESIS 20:854-859). Accordingly, removal of these proteins alone permits easier evaluation of less abundant proteins as disease markers. Accordingly, it is preferable to remove both immunoglobulin and albumin from the serum to produce an immunoglobulin and albumin depleted serum suitable for further processing.

The immunoglobulin and/or albumin proteins can be extracted using conventional methodologies, for example, affinity-based methodologies, known and used in the art. For example, immunoglobulin can be removed selectively from samples using binding proteins, for example, an antibody or a fragment thereof, Protein A, or Protein G, immobilized on a solid support. For example, a solution of interest can be passed through a chromatography column packed with such a solid support under conditions such that the immunoglobulin molecules preferentially bind to the matrix. The resulting column flow through, therefore, is depleted of immunoglobulin. A preferred matrix comprises Protein G coupled to agarose particles, available commercially from Pharmacia and Upjohn, Peapack, NJ under the trade name Hitrap Protein G. Similarly, albumin can be removed selectively for samples of interest via affinity chromatography, using, for example, Sepharose coupled to Cibacron blue available commercially from Pharmacia and Upjohn, Peapack, NJ. Alternatively, both albumin and immunoglobulin G can be removed simultaneously from serum using ProtoClear™ (Lollo *et al.* (1999) ELECTROPHORESIS 20:854-859). The authors report that greater than 95% of human serum albumin and greater than 97% of human immunoglobulin can be removed using ProtoClear™.

After depleting the samples of at least one abundant protein, the resulting sample then is fractionated to give a plurality of fractions, with each fraction comprising a plurality of molecules. The initial fractionation preferably is by a non-electrophoretic method, for example, by chromatography, more specifically, affinity chromatography. In a more preferred embodiment, the affinity chromatography is ion exchange chromatography, for example, anion or cation exchange chromatography. With serum, this step preferably is performed by anion exchange chromatography. During ion exchange chromatography, the sample of interest is combined with an appropriate matrix, for example, an anionic exchange matrix, and molecules are allowed to bind to the matrix. After washing to remove unbound material, the bound molecules then are eluted selectively into different elution buffers, each buffer preferentially eluting a different population of molecules. It is contemplated that by choosing appropriate buffers it is possible to generate a plurality of fractions, each comprising a plurality of molecules. In a procedure described in detail in Example 1, serum substantially free of immunoglobulin and albumin was subdivided into twelve fractions containing approximately equal amounts of protein by anion exchange chromatography. "Substantially free" is understood to mean at least 70%, more preferably at least 80%, more preferably at least 90% and most preferably at least 95% of a particular molecule. Anion exchange chromatography produces different populations of

samples, with each sample containing numerous molecules, but significantly less in number than the original starting material. These molecules can then be more easily characterized as a function of mass. In an exemplary protocol, serum is applied to a Mono Q (Pharmacia and Upjohn, Peapack, NJ) anion exchange column in phosphate buffer. The proteins once bound can be eluted by increasing the concentration of a salt, for example, sodium chloride, in a series of elution buffers. The choice of appropriate salt concentrations is considered to be within the level of skill in the art and will depend upon variables such as the type of starting material, and the types and numbers of proteins desired in each population.

Alternatively, the affinity chromatography may be performed using a solid support having carbohydrate binding moieties, for example, lectin, disposed thereon. As a result, it is possible to separate glycosylated from non-glycosylated molecules.

One or more of the resulting fractions can then be analyzed by mass, for example, mass-spectroscopy. For example, each fraction can be analyzed by matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectroscopy or by surface enhanced laser desorption/ionization-time of flight (SELDI-TOF) mass spectroscopy. See U.S. Patent No. 5,719,060.

Generally, analysis by mass spectrometry involves the vaporization and ionization of a sample of material using a high energy source, for example, a laser. Usually, the material is vaporized from the surface of a probe tip into the gas or vapor phase by a laser beam, whereby some of the individual molecules become ionized. The positively charged molecules then are accelerated using a high voltage field and allowed to fly into a high vacuum chamber, at the end of which is an detection surface. Because the time-of-flight is a function of mass of the ionized molecule, the elapsed time between ionization and impact can be used to determine molecule's mass. As a result, using this type of mass spectrometry it is possible to produce a profile of masses within the sample. By comparing the molecules present at a higher concentration in a sample from a mammal with the disease relative to those present in a sample from a mammal without the disease, it is possible to identify the molecules (i.e., markers) that are found at elevated levels in the diseased mammal.

Using mass spectrometry, it is further possible to characterize the markers by their binding affinity to a particular surface. For example, in SELDI-TOF mass spectroscopy, several different surfaces are available commercially from Ciphergen Biosystems, Inc., Palo Alto, CA.

Each of the surfaces have different surface properties and thus bind different populations of markers. Available surfaces include copper-treated surfaces and nickel-treated surfaces which can be generated by adding a copper or nickel salt solution to a chip comprising ethylenediaminetriacetic acid. Other SELDI chip surfaces include: WCX-2 which comprises carboxylate moieties, and SAX-2 which comprises quarternary ammonium moieties. The markers therefore can be further characterized by their affinity to a particular SELDI chip. For example, as used herein, the term "affinity" to a particular SELDI chip is understood to mean that a marker binds preferentially to one type of SELDI chip (*e.g.*, copper SELDI chip) relative to one or more of the other SELDI chips (*e.g.*, the nickel, SAX-2 and WCX-2 chips) disclosed herein. As discussed in detail in Example 1, comparison of the sera from diseased and healthy individuals revealed a number of proteins frequently present at detectable levels in the sera of diseased individuals, but infrequently present at comparable levels in the sera of healthy individuals.

Once the markers, for example, protein markers, have been identified by mass spectrometry, the identified proteins can be isolated by standard protein isolation methodologies and sequenced using protein sequencing technologies known and used in the art. For example, each of the markers, once identified, can be purified to homogeneity using the methodologies and the information derived therefrom in the previous steps. For example, the marker can be isolated based on its mass as determined by mass spectrometry and its other physical and chemical features, for example, ability to bind to an affinity column, for example, an ion exchange column. The proteins can be further characterized by conventional amino acid sequencing, for example, by Edman degradation and/or mass spectrometry-based microsequencing of proteolytic fragments.

It is contemplated that the method of the invention is particularly effective at identifying markers when the disease is cancer. Accordingly, it is contemplated that the method can be used to identify markers for breast cancer, lung cancer, prostate cancer, bladder cancer, cervical cancer, ovarian cancer, colon cancer or colorectal cancer. The Examples hereinbelow disclose the identification of breast cancer markers.

2. Detection of Disease

Once a disease marker has been identified, the marker, for example, a protein or a nucleic acid encoding the protein, may be used to determine whether an individual has the disease, and, if so, suitable detection methods can be used to monitor the status of the disease.

By using proteins or nucleic acids encoding the proteins as markers, the skilled artisan can produce a variety of detection methods for detecting a disease in a human. The methods typically comprise the steps of detecting, by some means, the presence of one or more markers in a tissue or body fluid sample of the human. The accuracy and/or reliability of the method for detecting markers in a human may be further enhanced by detecting the presence of a plurality of marker proteins or nucleic acids in a preselected tissue or body fluid sample. The detection assays may comprise one or more of the protocols described hereinbelow.

2.A. Protein-Based Assays

If the marker is a protein, the protein may be detected, for example, by combining the marker protein with a binding moiety capable of specifically binding the marker protein. The binding moiety may comprise, for example, a member of a ligand-receptor pair, i.e., a pair of molecules capable of having a specific binding interaction. The binding moiety may comprise, for example, a member of a specific binding pair, such as antibody-antigen, enzyme-substrate, nucleic acid-nucleic acid, protein-nucleic acid, protein-protein, or other specific binding pair known in the art. Binding proteins may be designed which have enhanced affinity for a target protein. Optionally, the binding moiety may be linked with a detectable label, such as an enzymatic, fluorescent, radioactive, phosphorescent or colored particle label. The labeled complex may be detected, e.g., visually or with the aid of a spectrophotometer or other detector.

Marker proteins may also be detected using gel electrophoresis techniques available in the art. In two-dimensional gel electrophoresis, the proteins are separated first in a pH gradient gel according to their isoelectric point. The resulting gel then is placed on a second polyacrylamide gel, and the proteins separated according to molecular weight (see, for example, O'Farrell (1975) *J. Biol. Chem.* 250: 4007-4021).

One or more marker proteins may be detected by first isolating proteins from a sample obtained from an individual suspected of having a disease, and then separating the proteins by two-dimensional gel electrophoresis to produce a characteristic two-dimensional gel

electrophoresis pattern. The pattern may then be compared with a standard gel pattern produced by separating, under the same or similar conditions, proteins isolated from normal or known cancer cells. The standard gel pattern may be stored in, and retrieved from an electronic database of electrophoresis patterns. The presence of a marker protein in the two-dimensional gel provides an indication that the sample being tested was taken from a person with the disease. As with the other detection assays described herein, the detection of two or more proteins, for example, in the two-dimensional gel electrophoresis pattern further enhances the accuracy of the assay. The presence of a plurality, e.g., two to five, marker proteins on the two-dimensional gel provides an even stronger indication of the presence of disease in the individual. The assay thus permits the early detection and treatment of the disease.

A marker protein may also be detected using any one of a wide range of immunoassay techniques available in the art. For example, the skilled artisan may employ a sandwich immunoassay format to detect a disease marker in a body fluid sample. Alternatively, the skilled artisan may use conventional immuno-histochemical procedures for detecting the presence of the marker in a tissue sample using one or more labeled binding proteins.

In a sandwich immunoassay, two antibodies capable of binding the marker protein generally are used, e.g., one immobilized onto a solid support, and one free in solution and labeled with a detectable chemical compound. Examples of chemical labels that may be used for the second antibody include radioisotopes, fluorescent compounds, and enzymes or other molecules that generate colored or electrochemically active products when exposed to a reactant or enzyme substrate. When a sample containing the marker protein is placed in this system, the marker protein binds to both the immobilized antibody and the labeled antibody, to form a "sandwich" immune complex on the support's surface. The complexed protein then is detected by washing away non-bound sample components and excess labeled antibody, and measuring the amount of labeled antibody complexed to protein on the support's surface. Alternatively, the antibody free in solution, which can be labeled with a chemical moiety, for example, a hapten, may be detected by a third antibody labeled with a detectable moiety which binds the free antibody or, for example, the hapten coupled thereto.

Both the sandwich immunoassay and the tissue immunohistochemical procedure are highly specific and very sensitive, provided that labels with good limits of detection are used. A detailed review of immunological assay design, theory and protocols can be found in numerous

texts in the art, including "*Practical Immunology*", Butt, W.R., ed., (1984) Marcel Dekker, New York and "*Antibodies, A Laboratory Approach*", Harlow *et al.* eds. (1988) Cold Spring Harbor Laboratory.

In general, immunoassay design considerations include preparation of antibodies (e.g., monoclonal or polyclonal antibodies) having sufficiently high binding specificity for the target protein to form a complex that can be distinguished reliably from products of nonspecific interactions. As used herein, the term "antibody" is understood to mean binding proteins, for example, antibodies or other proteins comprising an immunoglobulin variable region-like binding domain, having the appropriate binding affinities and specificities for the target protein. The higher the antibody binding specificity, the lower the target protein concentration that can be detected. As used herein, the terms "specific binding" or "binding specifically" are understood to mean that the binding moiety, for example, a binding protein has a binding affinity for the target protein of greater than about 10^5 M^{-1}, more preferably greater than about 10^7 M^{-1}.

Antibodies to an isolated marker or target protein which are useful in assays for detecting a breast cancer in an individual may be generated using standard immunological procedures well known and described in the art. See, for example, *Practical Immunology,* Butt, N.R., ed., Marcel Dekker, NY, 1984. Briefly, an isolated target protein is used to raise antibodies in a xenogeneic host, such as a mouse, goat or other suitable mammal. The marker protein is combined with a suitable adjuvant capable of enhancing antibody production in the host, and is injected into the host, for example, by intraperitoneal administration. Any adjuvant suitable for stimulating the host's immune response may be used. A commonly used adjuvant is Freund's complete adjuvant (an emulsion comprising killed and dried microbial cells and available from, for example, Calbiochem Corp., San Diego, or Gibco, Grand Island, NY). Where multiple antigen injections are desired, the subsequent injections may comprise the antigen in combination with an incomplete adjuvant (e.g., cell-free emulsion). Polyclonal antibodies may be isolated from the antibody-producing host by extracting serum containing antibodies to the protein of interest. Monoclonal antibodies may be produced by isolating host cells that produce the desired antibody, fusing these cells with myeloma cells using standard procedures known in the immunology art, and screening for hybrid cells (hybridomas) that react specifically with the target protein and have the desired binding affinity.

Antibody binding domains also may be produced biosynthetically and the amino acid sequence of the binding domain manipulated to enhance binding affinity with a preferred epitope on the target protein. Specific antibody methodologies are well understood and described in the literature. A more detailed description of their preparation can be found, for example, in Butt (1984) "*Practical Immunology*" (*supra*).

In addition, genetically engineered biosynthetic antibody binding sites, also known in the art as BABS or sFv's, may be used in the practice of the instant invention. Methods for making and using BABS comprising (i) non-covalently associated or disulfide bonded synthetic V_H and V_L dimers, (ii) covalently linked V_H-V_L single chain binding sites, (iii) individual V_H or V_L domains, or (iv) single chain antibody binding sites are disclosed, for example, in U.S. Patent Nos.: 5,091,513; 5,132,405; 4,704,692; and 4,946,778. Furthermore, BABS having requisite specificity for the marker protein can be derived by phage antibody cloning from combinatorial gene libraries (see, for example, Clackson *et al.* (1991) *Nature* 352: 624-628). Briefly, phage each expressing on their coat surfaces, BABS having immunoglobulin variable regions encoded by variable region gene sequences derived from mice pre-immunized with isolated marker proteins, or fragments thereof, are screened for binding activity against immobilized breast cancer-associated protein. Phage which bind to the immobilized marker proteins are harvested and the gene encoding the BABS sequenced. The resulting nucleic acid sequences encoding the BABS of interest may then be expressed in conventional expression systems to produce the BABS protein.

The isolated marker protein also may be used for the development of diagnostic and other tissue evaluating kits and assays to monitor the level of the proteins in a tissue or fluid sample. For example, the kit may include antibodies or other specific binding proteins which bind specifically to the marker proteins and which permit the presence and/or concentration of the marker proteins to be detected and/or quantitated in a tissue or fluid sample.

Suitable kits for detecting marker proteins are contemplated to include, *e.g.*, a receptacle or other means for capturing a sample to be evaluated, and means for detecting the presence and/or quantity in the sample of one or more of the marker proteins described herein. As used herein, "means for detecting" in one embodiment includes one or more antibodies specific for these proteins and means for detecting the binding of the antibodies to these proteins by, *e.g.*, a standard sandwich immunoassay as described herein. Where the presence of a protein within a

cell is to be detected, *e.g.*, as from a tissue sample, the kit also may comprise means for disrupting the cell structure so as to expose intracellular proteins.

2.B. Nucleic Acid-based Assays

The presence of a disease in an individual may also be determined by detecting, in a tissue or body fluid sample, a nucleic acid molecule encoding the marker protein. Using methods well known to those of ordinary skill in the art, the marker proteins may be sequenced, and then, based on the determined sequence, oligonucleotide probes designed for screening a cDNA library (see, for example, Sambrook *et al.* (1989) *supra*).

A target nucleic acid molecule encoding a marker protein may be detected using a labeled binding moiety capable of specifically binding the target nucleic acid. The binding moiety may comprise, for example, a protein, a nucleic acid or a peptidyl nucleic acid. Additionally, a target nucleic acid, such as an mRNA encoding a marker protein, may be detected by conducting, for example, a Northern blot analysis using labeled oligonucleotides, *e.g.*, nucleic acid fragments complementary to and capable of hybridizing specifically with at least a portion of a target nucleic acid.

More specifically, gene probes comprising complementary RNA or, preferably, DNA to the disease-associated nucleotide sequences or mRNA sequences encoding the marker proteins may be produced using established recombinant techniques or oligonucleotide synthesis. The probes hybridize with complementary nucleic acid sequences presented in the test specimen, and can provide exquisite specificity. A short, well-defined probe, coding for a single unique sequence is most precise and preferred. Larger probes generally are less specific. While an oligonucleotide of any length may hybridize to an mRNA transcript, oligonucleotides typically within the range of 8-100 nucleotides, preferably within the range of 15-50 nucleotides, are envisioned to be most useful in standard hybridization assays. Choices of probe length and sequence allow one to choose the degree of specificity desired. Hybridization is carried out at from 50° to 65°C in a high salt buffer solution, formamide or other agents to set the degree of complementarity required. The state of the art is such that probes can be manufactured to recognize essentially any DNA or RNA sequence. For further particulars, see, for example, *Guide to Molecular Techniques*, Berger *et al.*, Methods of Enzymology, Vol. 152, 1987.

A wide variety of different labels coupled to the probes or antibodies may be employed in the assays. The labeled reagents may be provided in solution or coupled to an insoluble support,

depending on the design of the assay. The various conjugates may be joined covalently or noncovalently, directly or indirectly. When bonded covalently, the particular linkage group will depend upon the nature of the two moieties to be bonded. A large number of linking groups and methods for linking are taught in the literature. Broadly, the labels may be divided into the following categories: chromogens; catalyzed reactions; chemiluminescence; radioactive labels; and colloidal-sized colored particles. The chromogens include compounds which absorb light in a distinctive range so that a color may be observed, or emit light when irradiated with light of a particular wavelength or wavelength range, *e.g.*, fluorescers. Both enzymatic and nonenzymatic catalysts may be employed. In choosing an enzyme, there will be many considerations including the stability of the enzyme, whether it is normally present in samples of the type for which the assay is designed, the nature of the substrate, and the effect if any of conjugation on the enzyme's properties. Potentially useful enzyme labels include oxiodoreductases, transferases, hydrolases, lyases, isomerases, ligases, or synthetases. Interrelated enzyme systems may also be used. A chemiluminescent label involves a compound that becomes electronically excited by a chemical reaction and may then emit light that serves as a detectable signal or donates energy to a fluorescent acceptor. Radioactive labels include various radioisotopes found in common use such as the unstable forms of hydrogen, iodine, phosphorus or the like. Colloidal-sized colored particles involve material such as colloidal gold that, in aggregate, form a visually detectable distinctive spot corresponding to the site of a substance to be detected. Additional information on labeling technology is disclosed, for example, in U.S. Patent No. 4,366,241.

A common method of *in vitro* labeling of nucleotide probes involves nick translation wherein the unlabeled DNA probe is nicked with an endonuclease to produce free 3'hydroxyl termini within either strand of the double-stranded fragment. Simultaneously, an exonuclease removes the nucleotide residue from the 5'phosphoryl side of the nick. The sequence of replacement nucleotides is determined by the sequence of the opposite strand of the duplex. Thus, if labeled nucleotides are supplied, DNA polymerase will fill in the nick with the labeled nucleotides. Using this well-known technique, up to 50% of the molecule can be labeled. For smaller probes, known methods involving 3'end labeling may be used. Furthermore, there are currently commercially available methods of labeling DNA with fluorescent molecules, catalysts, enzymes, or chemiluminescent materials. Biotin labeling kits are commercially available (Enzo Biochem Inc.) under the trademark Bio-Probe. This type of system permits the probe to be coupled to avidin with in turn is labeled with, for example, a fluorescent molecule, enzyme,

antibody, etc. For further disclosure regarding probe construction and technology, see, for example, Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor, N.Y., 1982).

The oligonucleotide selected for hybridizing to the target nucleic acid, whether synthesized chemically or by recombinant DNA methodologies, is isolated and purified using standard techniques and then preferably labeled (e.g., with ^{35}S or ^{32}P) using standard labeling protocols. A sample containing the target nucleic acid then is run on an electrophoresis gel, the dispersed nucleic acids transferred to a nitrocellulose filter and the labeled oligonucleotide exposed to the filter under stringent hybridizing conditions, *e.g.* 50% formamide, 5 X SSPE, 2 X Denhardt's solution, 0.1% SDS at 42°C, as described in Sambrook *et al.* (1989) *supra.* The filter may then be washed using 2 X SSPE, 0.1% SDS at 68°C, and more preferably using 0.1 X SSPE, 0.1% SDS at 68°C. Other useful procedures known in the art include solution hybridization, and dot and slot RNA hybridization. Optionally, the amount of the target nucleic acid present in a sample then is quantitated by measuring the radioactivity of hybridized fragments, using standard procedures known in the art.

In addition, oligonucleotides may also be used to identify other sequences encoding members of the target protein families. The methodology may also be used to identify genetic sequences associated with the nucleic acid sequences encoding the proteins described herein, *e.g.*, to identify non-coding sequences lying upstream or downstream of the protein coding sequence, and which may play a functional role in expression of these genes. Additionally, binding assays may be conducted to identify and detect proteins capable of a specific binding interaction with a nucleic acid encoding a breast cancer-associated protein, which may be involved, *e.g.*, in gene regulation or gene expression of the protein. In a further embodiment, the assays described herein may be used to identify and detect nucleic acid molecules comprising a sequence capable of recognizing and being specifically bound by a marker protein.

In addition, it is anticipated that using a combination of appropriate oligonucleotide primers, *i.e.*, more than one primer, the skilled artisan may determine the level of expression of a target gene *in vivo* by standard polymerase chain reaction (PCR) procedures, for example, by quantitative PCR. Conventional PCR based assays are discussed, for example, in Innes *et al* (1990) "*PCR Protocols; A guide to methods and Applications*", Academic Press and Innes *et al.* (1995) "*PCR Strategies*" Academic Press, San Diego, CA.

Recombinant marker molecules can be produced as described hereinbelow. For example, DNA encoding the marker molecules can be inserted, using conventional techniques well described in the art (see, for example, Sambrook (1989) *supra*) into any of a variety of expression vectors and transfected into an appropriate host cell to produce recombinant proteins, including both full length and truncated forms. Useful host cells include *E. coli, Saccharomyces cerevisiae, Pichia pastoris,* the insect/baculovirus cell system, myeloma cells, and various other mammalian cells. The full length forms of such proteins are preferably expressed in mammalian cells, as disclosed herein. The vector can additionally include various sequences to promote correct expression of the recombinant protein, including transcription promoter and termination sequences, enhancer sequences, preferred ribosome binding site sequences, preferred mRNA leader sequences, preferred protein processing sequences, preferred signal sequences for protein secretion, and the like. The DNA sequence encoding the gene of interest can also be manipulated to remove potentially inhibiting sequences or to minimize unwanted secondary structure formation. As will be appreciated by the practitioner in the art, the recombinant protein can also be expressed as a fusion protein.

After translation, the protein can be purified from the cells themselves or recovered from the culture medium. The DNA can also include sequences which aid in expression and/or purification of the recombinant protein. The DNA can be expressed directly or can be expressed as part of a fusion protein having a readily cleavable fusion junction.

In one preferred embodiment, the DNA is expressed in a suitable mammalian host. Useful hosts include fibroblast 3T3 cells, (e.g., NIH 3T3, from CRL 1658) COS (simian kidney ATCC, CRL-1650) or CHO (Chinese hamster ovary) cells (e.g., CHO-DXB11, from Chasin (1980) *Proc. Nat'l. Acad. Sci. USA* 77 :4216-4222), mink-lung epithelial cells (MV1Lu), human foreskin fibroblast cells, human glioblastoma cells, and teratocarcinoma cells. Other useful eukaryotic cell systems include yeast cells, the insect/baculovirus system or myeloma cells.

In order to express a marker protein molecule, the DNA is subcloned into an insertion site of a suitable, commercially available vector along with suitable promoter/enhancer sequences and 3' termination sequences. Useful promoter/enhancer sequence combinations include the CMV promoter (human cytomegalovirus (MIE) promoter) present, for example, on pCDM8, as well as the mammary tumor virus promoter (MMTV) boosted by the Rous sarcoma virus LTR enhancer sequence (e.g., from Clontech, Inc., Palo Alto). A useful inducible promoter includes,

for example, a Zn^{2+}-inducible promoter, such as the Zn^{2+} metallothionein promoter (Wrana *et al.* (1992) Cell 71: 1003-1014). Other inducible promoters are well known in the art and can be used with similar success. Expression also can be further enhanced using *trans*-activating enhancer sequences. The plasmid also preferably contains an amplifiable marker, such as DHFR under suitable promoter control, e.g., SV40 early promoter (ATCC #37148). Transfection, cell culturing, gene amplification and protein expression conditions are standard conditions, well known in the art, such as are described, for example in Ausubel *et al.*, ed., (1989) *"Current Protocols in Molecular Biology"*, John Wiley & Sons, NY. Briefly, transfected cells are cultured in medium containing 5-10% dialyzed fetal calf serum (dFCS), and stably transfected high expression cell lines obtained by amplification and subcloning and evaluated by standard Western and Northern blot analysis. Southern blots also can be used to assess the state of integrated sequences and the extent of their copy number amplification.

The expressed candidate protein is then purified using standard procedures. A currently preferred methodology uses an affinity column, such as a ligand affinity column or an antibody affinity column. The column then is washed, and the candidate molecules selectively eluted in a gradient of increasing ionic strength, changes in pH, or addition of mild detergent. It is appreciated that in addition to the candidate molecules which bind to the breast cancer-associated proteins, the breast cancer associated proteins themselves may likewise be produced using such recombinant DNA technologies.

The following non-limiting examples provide details for the isolation and characterization of breast cancer markers together with methods of using the markers for the detection of breast cancer. It is contemplated that the same or a similar protocol can be used to identify markers for other diseases, for example, other cancers.

Example 1 – Identification of Breast Cancer Markers

To identify markers for breast cancer, the sera of individuals with breast cancer were compared to the sera of normal individuals using the following protocol. Briefly, 0.5 mL aliquots of sera harvested from the individuals were thawed. Then, 1 μL of a 1 mg/mL solution of soybean trypsin inhibitor (SBTI) and 1 μL of a 1 mg/mL solution of leupeptin were added to each aliquot. To remove lipids, 350 μL of 1,1,2-trifluorotrichloroethane was added to each sample. The samples then were vortexed for five minutes and centrifuged in a microcentrifuge for five minutes at 4°C. The resulting supernatants were applied to a 1 mL column of agarose

coupled to protein G (Hitrap Protein G column, Pharmacia and Upjohn, Peapack, NJ) to remove immunoglobulin proteins. The column then was rinsed with 3 mL of 50 mM sodium phosphate, pH 7.0, with SBTI and leupeptin ("binding buffer"), and the resulting flowthrough applied directly to a 5 mL column of 6% Sepharose coupled to Cibacron blue (Hitrap blue column, Pharmacia and Upjohn, Peapack, NJ) to remove albumin proteins. The Hitrap blue column was rinsed with 20 mL of binding buffer. The resulting flow through was concentrated using four centrifugation-based concentrators with a 10kD cutoff (Centricon 10, Millipore Corporation, Bedford, MA) to give a final volume of about 0.7 mL.

The resulting serum (substantially free of immunoglobulin and albumin) was subdivided into twelve fractions containing approximately equal amounts of protein by anion exchange chromatography. Specifically, the serum was applied to a Mono Q (Pharmacia and Upjohn, Peapack, NJ) anion exchange column (a strong anion exchanger with quarternary ammonium groups) in 50 mM sodium phosphate buffer, pH 7.0 and proteins were eluted from the column by increasing the concentration of sodium chloride in a stepwise manner. In this protocol, the serum was divided into twelve fractions based on the concentration of sodium chloride used for elution. These fractions accordingly were designated flow through, 25 mM, 50 mM, 75 mM, 100 mM, 125 mM, 150 mM, 200 mM, 250 mM, 300 mM, 400 mM, and 2M sodium chloride. After elution, each fraction was concentrated to approximately 100 μg/mL and buffer exchanged into binding buffer.

Then 4-10 μL from each of the twelve fractions were applied and allowed to bind to each of four SELDI chip surfaces, each surface holding up to eight samples. The intended location of each sample on the chip was demarcated with a circle drawn using a hydrophobic marker like those used in Pap smears. The SELDI chips used herein were purchased from Ciphergen Biosystems, Inc., Palo Alto, California, and used as described below.

For copper or nickel surfaces, a chip containing ethylenediaminetriacetic acid moieties (IMAC, Ciphergen Biosystems, Inc., Palo Alto, CA) was pretreated with two five-minute applications of five μL of a copper salt or nickel salt solution, and washed with deionized water. After a five-minute treatment with five μL of binding buffer, two to three microliters of sample were applied to the surface for thirty to sixty minutes. Another two to three microliters of sample then were applied for an additional thirty to sixty minutes. The chips then were washed twice with binding buffer to remove unbound proteins. 0.5 μL of sinapinic acid (12.5 mg/mL) was

added twice and allowed to dry each time. The presence of sinapinic acid enhances the vaporization and ionization of the bound proteins upon mass spectrometry.

For chip surfaces containing carboxyl moieties (WCX-2, Ciphergen Biosystems, Inc., Palo Alto, CA), before use of the hydrophobic pen, the surface was washed with 10 mM HCl for thirty minutes and rinsed five times with deionized water. After use of the pen, the surface was washed five times with 5 μL of binding buffer and once with deionized water. Two to three μL of sample were applied in two applications of thirty to sixty minutes each. The surface was washed twice with 5 μL of binding buffer, and 0.5 μL of sinapinic acid were applied twice.

For chip surfaces containing quarternary ammonium moieties (SAX-2, Ciphergen Biosystems, Inc., Palo Alto, CA), after use of the pen, the surface was washed five times with five μL of binding buffer and once with deionized water. Application of sample, washing, and application of sinapinic acid were performed as described above.

The chips then were subjected to mass spectrometry utilizing a Ciphergen SELDI PBS One (Ciphergen Biosystems, Inc., Palo Alto, CA) running the software program "SELDI v. 2.0". For all chips, "high mass" was set to 200,000 Daltons, "starting detector sensitivity" was set to 9 (from a range of 1-10, with 10 being the highest sensitivity), NDF (neutral density filter) was set to "OUT", data acquisition method was set to "Seldi Quantitation", SELDI acquisition parameters were set to 20, with increments of 5, and warming with two shots at intensity 50 (out of 100) was included. For IMAC chips, mass was optimized from 3000 Daltons to 3001 Daltons, starting laser intensity was set to 80 (out of 100), and transients set to 5 (i.e., 5 laser shots per site). Peaks were identified automatically by computer. For WCX-2 chips, mass was optimized from 3,000 Daltons to 50,000 Daltons, starting laser intensity was set to 80, and transients set to 8. Peaks were identified automatically by computer. For SAX-2 chips, mass was optimized from 3,000 Daltons to 50,000 Daltons, starting laser intensity was set to 85, and transients set to 8. Peaks were identified automatically by computer.

Ten serum samples (five from normal individuals and five from individuals with breast cancer) were analyzed by mass spectrometry to identify the proteins present in the sixty fractions described above. The resulting peaks in the mass spectrometry trace were compared to identify those peaks present in the serum samples from individuals with breast cancer but not present in the normal samples. If peaks in different samples had a mass difference of no more than one percent, the peaks were presumed to be the same. Eleven mass spectrometry peaks ranging in

size from just over 11,000 Da to approximately 103,000 Da were identified as present in all five serum samples from individuals with breast cancer and in none of the samples from normal individuals. The presence or absence of these peaks then was determined for an additional thirty serum samples (fifteen from normal individuals and fifteen from individuals with breast cancer). Seven other peaks that were present in four of the original five breast cancer serum samples, but not in any of the normal samples, were also analyzed because they were present in the same fraction and on the same SELDI surface as one or more of the eleven peaks already under evaluation. Of the eighteen peaks studied, fifteen were present in fifteen or more of the twenty breast cancer serum samples, but absent from 15 or more of the normal serum samples.

The results of the foregoing analyses are summarized in Table 1. The masses listed in the table are presumed accurate to within one percent.

TABLE 1.

Mass (Da)	Mono Q fraction (mM sodium chloride)	SELDI chip surface used	Number of positive samples from individuals with breast cancer	Number of positive samples from individuals without breast cancer
16210	0 (flow-through)	Nickel	17	1
17188	25 mM	WCX-2	17	2
30183	25 mM	WCX-2	15	3
34664	25 mM	WCX-2	16	4
20050	50 mM	Nickel	19	0
28258	50 mM	Nickel	20	0
24170	50 mM	Nickel	17	0
35393	50 mM	Nickel	17	3
34908	50 mM	WCX-2	16	2
70908	100 mM	WCX-2	20	0
17840	100 mM	WCX-2	18	2
11709	150 mM	SAX-2	20	0
42354	200 mM	Nickel	17	0
56280	200 mM	Nickel	16	0
34517	400 mM	Copper	18	1

Example 2 – Purification and Characterization of 28.3 kD Breast Cancer Protein

Breast cancer-associated proteins based upon the biochemical and mass spectrometry data provided above may be better characterized using well-known techniques. For example, samples of the serum can be fractionated using, for example, column chromatography and/or

electrophoresis, to produce purified protein samples corresponding to each of the proteins identified in Table 1. The sequences of the isolated proteins can then be determined using conventional peptide sequencing methodologies. It is appreciated that the skilled artisan, in view of the foregoing disclosure, would be able to produce an antibody directed against any breast cancer-associated protein identified by the methods described herein. Moreover, the skilled artisan, in view of the foregoing disclosure, would be able to produce nucleic acid sequences that encode the fragments described above, as well as nucleic acid sequences complementary thereto. In addition, the skilled artisan using conventional recombinant DNA methodologies, for example, by screening a cDNA library with such a nucleic acid sequence, would be able to isolate full length nucleic acid sequences encoding target breast cancer-associated proteins. Such full length nucleic acid sequences, or fragments thereof, may be used to generate nucleic acid-based detection systems or therapeutics.

The 28.3 kD breast cancer protein identified in Example 1 was isolated and further characterized as follows. Approximately 30 mL of serum (combined from multiple breast cancer patients) was depleted of immunoglobulin G and serum albumin using Protein G chromatography and Cibacron Blue agarose chromatography, respectively, using standard methodologies such as those described in Example 1. The albumin and immunoglobulin depleted serum was then fractionated by Mono Q ion-exchange affinity chromatography. Briefly, the serum proteins were applied to a 5 mL Mono Q column (Pharmacia and Upjohn, Peapack, NJ) in 50mM sodium phosphate buffer, pH 7.0, and the flow through fraction collected. Thereafter, the serum proteins were eluted stepwise from the column using 50mM sodium phosphate buffer, pH 7.0 containing increasing concentrations of sodium chloride. In this manner, 12 serum fractions were obtained, each containing a different amount of sodium chloride. The fractions included flow through, and elution buffers of 50 mM sodium phosphate buffer, pH 7.0 containing 25mM, 50mM, 75mM, 100mM, 125mM, 150mM, 200mM, 250mM, 300mM, 400mM, and 2M sodium chloride.

The 50mM sodium chloride fraction containing the protein of interest was subsequently buffer exchanged back into 50mM sodium phosphate buffer, pH 7.0 and concentrated by means of a Centricon 10 (Millipore) in accordance with the manufacturers instructions. The resulting sample then was fractionated by size exclusion chromatography on a Sephacryl S-200 column (Pharmacia) using an isocratic buffer containing 100mM sodium phosphate, 150 mM NaCl, pH

7.4. Fractions that eluted from the column were evaluated for the presence of the 28.3kD protein using the Ciphergen SELDI mass spectroscopy as described in Example 1. Fractions containing the 28.3 kD protein were pooled and applied to an IMAC column (Sigma) which had been pre-loaded with Ni^{2+}, by prior incubation with 50mM $NiCl_2$. The IMAC column then was washed with 6 bed volumes of a solution containing 100mM sodium phosphate, 150 mM NaCl, pH 7.4, and the bound protein fraction eluted with the same solution containing 100mM imidazole. The eluted fraction then was concentrated by means of a Minicon 10 (Millipore) and then was fractionated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on a 12% Tris glycine SDS-PAGE gel. Samples of the protein fraction were applied to two separate lanes of the gel. After electrophoresis, the resulting gel then was stained with Coomassie Brilliant Blue dye and destained to reveal the presence of proteins. Three bands of about 28.3 kD (characterized as the heaviest molecular weight protein, the medium molecular weight protein, and the lightest molecular weight protein) were excised from one of the 2 lanes and were eluted from the acrylamide slices.

The proteins were eluted from the gel as follows. Briefly, the gel slices were washed five times with HPLC grade water with vigorous vortexing. The washed slices then were cut into small pieces in 120μL of 100mM sodium acetate pH 8.5, 0.1% SDS and incubated overnight at 37°C. The supernatant was decanted into a fresh tube and dried in a speedvac. The resulting pellet then was reconstituted in 37 μL HPLC grade water. Approximately 1480 μL of cold ethanol then was added and the resulting mixture incubated overnight at -20°C. Thereafter, the sample was centrifuged at 4°C for 15 minutes at 11,000 rpm. The supernatant was removed and the resulting pellet reconstituted in 5 μL of water. The resulting protein solutions were run on the SELDI and the 28.3kD protein was identified in one of the three preparations (see Fig. 1A which corresponds to the heaviest 28 kD protein). The corresponding band then was excised from the second of the 2 lanes on the gel. After proteolysis with trypsin, the tryptic fragments were eluted from the gel and submitted for microsequence analysis via mass spectrometry.

Four individual masses were detected by mass spectrometry. When the four masses were used to search the Swiss Protein Database, all four masses were found to match amino acid sequences present in the protein referred to in the art as U2 small nuclear ribonucleoprotein B" (U2 snRNP B") (Habets *et al.* (1987) *supra*, Swiss Protein Database Accession Number 4507123). The results are summarized in Table 2.

TABLE 2.

Peptide	Sequence	SEQ ID NO:	Protein
1	RQLQGFPFYGKPMRI	1	U2 snRNP B"
2	RHDIAFVEFENDGQAGAARD	2	U2 snRNP B"
3	RLVPGRHDIAFVEFENDGQAGAARD	3	U2 snRNP B"
4	TVEQTATTTNK	4	U2 snRNP B"

The amino acid sequence, in an N- to C- terminal direction, of the U2 SnRNP B" protein in single amino acid code is :

MDIRPNHTIY INNMNDKIKK EELKRSLYAL FSQFGHVVDI VALKTMKMRG QAFVIFKELG SSTNALRQLQ GFPFYGKPMR IQYAKTDSDI ISKMRGTFAD KEKKKEKKKA KTVEQTATTT NKKPGQGTPN SANTQGNSTP NPQVPDYPPN YILFLNNLPE ETNEMMLSML FNQFPGFKEV RLVPGRHDIA FVEFENDGQA GAARDALQGF KITPSHAMKI TYAKK (SEQ ID NO: 5).

The 28.3 kD has been identified to be U2 SnRNP B" and, thus, it is contemplated that it is possible to use this protein or a nucleic acid encoding this protein as a target in an assay for detecting the presence of breast cancer in an individual. The development of such assays, once the marker has been identified, is considered to be within the level of the art.

Example 3 - Production of Antibodies Which Bind Specifically to Breast Cancer-associated Proteins

Once identified, a breast cancer-associated protein may be detected in a tissue or body fluid sample using numerous binding assays that are well known to those of ordinary skill in the art. For example, as discussed above, a breast cancer-associated protein may be detected in either a tissue or body fluid sample using an antibody, for example, a monoclonal antibody, which binds specifically to an epitope disposed upon the breast cancer-associated protein. In such detection systems, the antibody preferably is labeled with a detectable moiety.

Provided below is an exemplary protocol for the production of an anti-breast cancer-associated monoclonal antibody. Other protocols also are envisioned. Accordingly, the

particular method of producing antibodies to target proteins is not envisioned to be an aspect of the invention.

Balb/c by J mice (Jackson Laboratory, Bar Harbor, ME) are injected intraperitoneally with the target protein every 2 weeks until the immunized mice obtain the appropriate serum titer. Thereafter, the mice are injected with 3 consecutive intravenous boosts. Freund's complete adjuvant (Gibco, Grand Island) is used in the first injection, incomplete Freund's in the second injection; and saline is used for subsequent intravenous injections. The animal then is sacrificed and its spleen removed. Spleen cells (or lymph node cells) then are fused with a mouse myeloma line, *e.g.*, using the method of Kohler *et al.* (1975) *Nature* 256: 495. Hybridomas producing antibodies that react with the target proteins then are cloned and grown as ascites. Hybridomas are screened by reactivity to the immunogen in any desirable assay. Detailed descriptions of screening protocols, ascites production and immunoassays also are disclosed in PCT/US92/09220 published May 13, 1993.

Example 4 - Antibody-based Assay for Detecting Breast Cancer in an Individual

The following assay has been developed for tissue samples; however, it is contemplated that similar assays for testing fluid samples may be developed without undue experimentation. A typical assay may employ a commercial immunodetection kit, for example, the ABC Elite Kit from Vector Laboratories, Inc.

A biopsy sample is removed from the patient under investigation in accordance with the appropriate medical guidelines. The sample then is applied to a glass microscope slide and the sample fixed in cold acetone for 10 minutes. Then, the slide is rinsed in distilled water and pretreated with a hydrogen peroxide containing solution (2 mL 30% H_2O_2 and 30 mL cold methanol). The slide then is rinsed in a Buffer A comprising Tris Buffered Saline (TBS) with 0.1% Tween and 0.1% Brij. A mouse anti-breast cancer-associated protein monoclonal antibody in Buffer A is added to the slide and the slide then incubated for one hour at room temperature. The slide then is washed with Buffer A, and a secondary antibody (ABC Elite Kit, Vector Labs, Inc) in Buffer A is added to the slide. The slide then is incubated for 15 minutes at 37°C in a humidity chamber. The slides are washed again with Buffer A, and the ABC reagent (ABC Elite Kit, Vector Labs, Inc.) is then added to the slide for amplification of the signal. The slide is then incubated for a further 15 minutes at 37°C in the humidity chamber.

The slide then is washed in distilled water, and a diaminobenzedine (DAB) substrate added to the slide for 4-5 minutes. The slide then is rinsed with distilled water, counterstained with hematoxylin, rinsed with 95% ethanol, rinsed with 100% ethanol, and then rinsed with xylene. A cover slip is then applied to the slide and the result observed by light microscopy.

Equivalents

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in all respects illustrative rather than limiting on the invention described herein. Scope of the invention is thus indicated by the appended claims rather than by the foregoing description, and all changes that come within the meaning and range of equivalency of the claims are intended to be embraced by reference therein.

Incorporation By Reference

The entire disclosure of each of the aforementioned patent and scientific documents cited hereinabove is expressly incorporated by reference herein.

What is claimed is:

1. A method for identifying a marker molecule indicative of a disease in a mammal, the method comprising the steps of:

(a) removing an abundant protein from a sample harvested from a mammal with the disease;

(b) fractionating the sample produced by step (a) to produce a plurality of fractions, each fraction comprising a plurality of molecules;

(c) separating molecules disposed within a fraction produced by step (b) by mass;

(d) repeating steps (a) through (c) with a sample harvested from a mammal without the disease; and

(e) comparing the molecules separated by step (c) with those separated by step (d) to identify a marker molecule present at a higher concentration in the sample from the mammal with the disease relative to the sample from the mammal without the disease, wherein the marker molecule is indicative of the disease.

2. The method of claim 1, wherein in step (a) the sample is a body fluid.

3. The method of claim 2, wherein the body fluid is blood, serum, plasma, sweat, tears, urine, peritoneal fluid, lymph, vaginal secretion, semen, spinal fluid, ascitic fluid, saliva, sputum, or breast exudate.

4. The method of claim 2, wherein the body fluid is serum.

5. The method of claim 1, wherein the abundant protein comprises greater than 5% (w/w) of total protein in the sample.

6. The method of claim 5, wherein the abundant protein comprises greater than 20% (w/w) of total protein in the sample.

7. The method of claim 5, wherein the abundant protein is an immunoglobulin or an albumin.

8. The method of claim 6, wherein the abundant protein is an immunoglobulin or an albumin.

9. The method of claim 1, wherein in step (b) the fractionation is by a non-electrophoretic method.

10. The method of claim 1, wherein in step (b) the fractionation is by affinity chromatography.

11. The method of claim 10, wherein the affinity chromatography is ion exchange chromatography.

12. The method of claim 11, wherein the ion exchange chromatography is anion exchange chromatography.

13. The method of claim 10, wherein the affinity chromatography employs a solid support having carbohydrate binding moieties disposed thereon.

14. The method of claim 13, wherein the carbohydrate binding moieties comprise lectin.

15. The method of claim 1, wherein in step (c) the molecules are separated by matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectroscopy or by surface enhanced laser desorption/ionization-time of flight (SELDI-TOF) mass spectroscopy.

16. The method of claim 1, wherein in step (e) the marker molecule is detectable in the sample from the mammal with the disease but is not detectable in the sample from the mammal without the disease.

17. The method of claim 1, wherein the marker molecule is a protein.

18. The method of claim 1, wherein the disease is cancer.

19. The method of claim 18, wherein the cancer is breast cancer, lung cancer, prostate cancer, bladder cancer, cervical cancer, ovarian cancer, colon cancer or colorectal cancer.

20. The method of claim 1, wherein the mammal is a human.

21. A method for identifying a marker molecule indicative of a disease in a mammal, the method comprising the steps of:

(a) removing an abundant protein from a body fluid sample harvested from a mammal with the disease;

(b) fractionating the sample produced by step (a) by ion exchange chromatography to produce a plurality of fractions, each fraction comprising a plurality of molecules;

(c) separating molecules disposed within a fraction produced by step (b) by surface enhanced laser desorption/ionization-time of flight mass spectroscopy;

(d) repeating steps (a) through (c) with a body fluid sample harvested from a mammal without the disease; and

(e) comparing the molecules produced by step (c) with those produced by step (d) to identify a marker molecule present at a higher concentration in the sample from the mammal with the disease relative to the sample from the mammal without the disease, wherein the marker molecule is indicative of the disease.

22. The method of claim 21, wherein in step (a) the body fluid sample is blood, serum, plasma, sweat, tears, urine, peritoneal fluid, lymph, vaginal secretions, semen, spinal fluid, ascitic fluid, saliva, sputum, or breast exudate.

23. The method of claim 21, wherein in step (a) the body fluid sample is serum.

24. The method of claim 21, wherein the abundant protein constitutes greater than 5% (w/w) of total protein in the sample.

25. The method of claim 24, wherein the abundant protein constitutes greater than 20% (w/w) of total protein in the sample.

26. The method of claim 24, wherein the abundant protein is an immunoglobulin or an albumin.

27. The method of claim 20, wherein the ion exchange chromatography is anion exchange chromatography.

28. The method of claim 21, wherein in step (e) the marker molecule is detectable in the body fluid sample from the mammal with the disease but is not detectable in the body fluid sample from the mammal without the disease.

29. The method of claim 21, wherein the marker molecule is a protein.

30. The method of claim 21, wherein the disease is cancer.

31. The method of claim 30, wherein the cancer is breast cancer, lung cancer, prostate cancer, bladder cancer, cervical cancer, ovarian cancer, colon cancer or colorectal cancer.

32. The method of claim 20, wherein the mammal is a human.

33. An isolated marker molecule identified by the method of claim 1.

34. An isolated marker molecule identified by the method of claim 32.

35. A method of diagnosing a disease in a mammal, the method comprising the steps of:

(a) contacting a sample from the mammal with a binding moiety that binds specifically to a disease-associated protein to produce a binding moiety-disease-associated protein complex, wherein the binding moiety binds specifically to a marker protein identified by the method of claim 17; and

(b) detecting the presence of the complex, which if present is indicative of the presence of disease in the mammal.

36. The method of claim 35, wherein the binding moiety is an antibody.

37. The method of claim 36, wherein the antibody is a monoclonal antibody.

38. The method of claim 36, wherein the antibody is a polyclonal antibody.

39. The method of claim 36, wherein the antibody is labeled with a detectable moiety.

40. The method of claim 39, wherein the detectable moiety comprises a label selected from the group consisting of a radioactive label, a hapten label, a fluorescent label, and an enzymatic label.

41. The method of claim 35, wherein the disease is cancer.

42. The method of claim 35, wherein the mammal is a human.

43. A method of diagnosing a disease in a mammal, the method comprising the steps of:

(c) contacting a sample from the mammal with a binding moiety that binds specifically to a disease-associated protein to produce a binding moiety-disease-associated protein complex, wherein the binding moiety binds specifically to a marker protein identified by the method of claim 29; and

(d) detecting the presence of the complex, which if present is indicative of the presence of disease in the mammal.

44. The method of claim 43, wherein the binding moiety is an antibody.

45. The method of claim 44, wherein the antibody is a monoclonal antibody.

46. The method of claim 44, wherein the antibody is a polyclonal antibody.

47. The method of claim 44, wherein the antibody is labeled with a detectable moiety.

48. The method of claim 47, wherein the detectable moiety comprises a label selected from the group consisting of a radioactive label, a hapten label, a fluorescent label, and an enzymatic label.

49. The method of claim 43, wherein the disease is cancer.

50. The method of claim 43, wherein the mammal is a human.

1/1

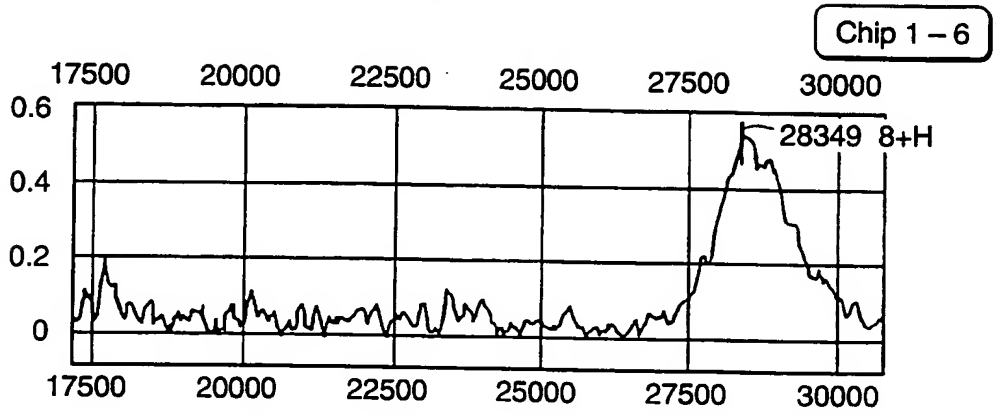

FIG. 1A

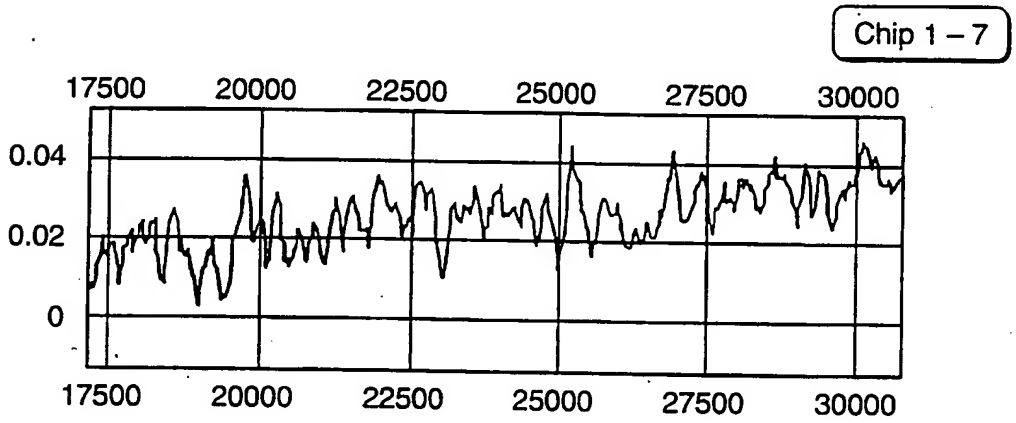

FIG. 1B

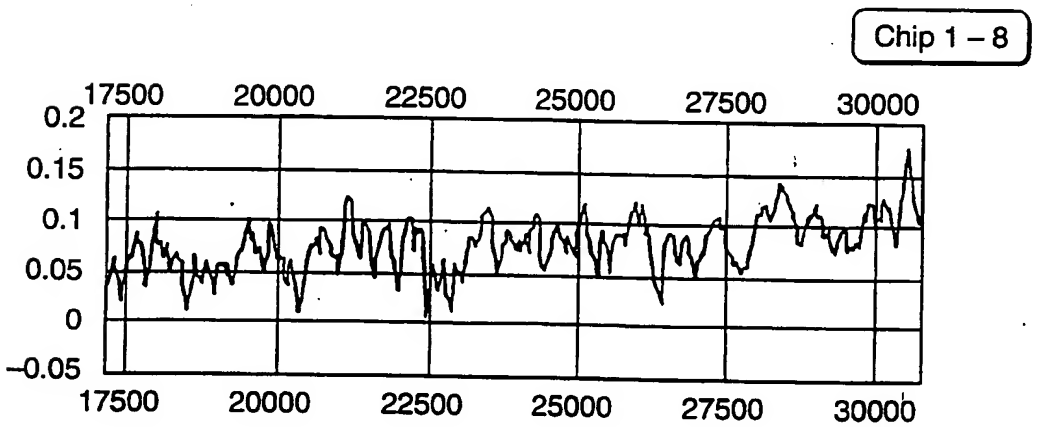

FIG. 1C

SEQUENCE LISTING

<110> Watkins, Brynmor

<120> Methods and Compositions for Identifying Disease Markers

<130> MTP-026PC

<140>
<141>

<150> US 60/165,173
<151> 1999-11-16

<150> US 60/172,170
<151> 1999-12-17

<150> US 60/178,860
<151> 2000-01-27

<150> US 60/201,721
<151> 2000-05-03

<160> 23

<170> PatentIn Ver. 2.0

<210> 1
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Tryptic peptide

<400> 1
Gln Leu Gln Gly Phe Pro Phe Tyr Gly Lys Pro Met Arg
1 5 10

<210> 2
<211> 18
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Tryptic peptide

<400> 2
His Asp Ile Ala Phe Val Glu Phe Glu Asn Asp Gly Gln Ala Gly Ala
1 5 10 15

Ala Arg

<210> 3
<211> 23
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Tryptic peptide

<400> 3
Leu Val Pro Gly Arg His Asp Ile Ala Phe Val Glu Phe Glu Asn Asp
1 5 10 15

Gly Gln Ala Gly Ala Ala Arg
20

<210> 4
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Tryptic peptide

<400> 4
Thr Val Glu Gln Thr Ala Thr Thr Thr Asn Lys
1 5 10

<210> 5
<211> 225
<212> PRT
<213> Homo sapiens

<400> 5
Met Asp Ile Arg Pro Asn His Thr Ile Tyr Ile Asn Asn Met Asn Asp
1 5 10 15

Lys Ile Lys Lys Glu Glu Leu Lys Arg Ser Leu Tyr Ala Leu Phe Ser
20 25 30

Gln Phe Gly His Val Val Asp Ile Val Ala Leu Lys Thr Met Lys Met
35 40 45

Arg Gly Gln Ala Phe Val Ile Phe Lys Glu Leu Gly Ser Ser Thr Asn
50 55 60

Ala Leu Arg Gln Leu Gln Gly Phe Pro Phe Tyr Gly Lys Pro Met Arg
65 70 75 80

Ile Gln Tyr Ala Lys Thr Asp Ser Asp Ile Ile Ser Lys Met Arg Gly
85 90 95

Thr Phe Ala Asp Lys Glu Lys Lys Lys Glu Lys Lys Lys Ala Lys Thr
100 105 110

Val Glu Gln Thr Ala Thr Thr Thr Asn Lys Lys Pro Gly Gln Gly Thr
115 120 125

Pro Asn Ser Ala Asn Thr Gln Gly Asn Ser Thr Pro Asn Pro Gln Val
130 135 140

Pro Asp Tyr Pro Pro Asn Tyr Ile Leu Phe Leu Asn Asn Leu Pro Glu
145 150 155 160

Glu Thr Asn Glu Met Met Leu Ser Met Leu Phe Asn Gln Phe Pro Gly
165 170 175

Phe Lys Glu Val Arg Leu Val Pro Gly Arg His Asp Ile Ala Phe Val
180 185 190

Glu Phe Glu Asn Asp Gly Gln Ala Gly Ala Ala Arg Asp Ala Leu Gln
195 200 205

Gly Phe Lys Ile Thr Pro Ser His Ala Met Lys Ile Thr Tyr Ala Lys
210 215 220

Lys
225

<210> 6
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 6
Gly Gln Val Pro Met Gln Asp Pro Arg
1 5

<210> 7
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 7
Gly Ser Leu Pro Ala Asn Val Pro Thr Pro Arg
1 5 10

<210> 8
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 8
Gly Leu Leu Gly Asp Ala Pro Asn Asp Pro Arg
1 5 10

<210> 9
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 9
Ala Gly Leu Thr Val Arg Asp Pro Ala Val Asp Arg
1 5 10

<210> 10
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 10
Ala Leu Arg Val Asp Asn Ala Ala Ser Glu Lys Asn Lys
1 5 10

<210> 11
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 11
Gly Gly Thr Leu Leu Ser Val Thr Gly Glu Val Glu Pro Arg
1 5 10

<210> 12
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 12
Asp Ile Phe Ser Glu Val Gly Pro Val Val Ser Phe Arg
1 5 10

<210> 13
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 13
Gly Ile Asp Ala Arg Gly Met Glu Ala Arg Ala Met Glu Ala Arg
1 5 10 15

<210> 14
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 14
Gly Met Glu Ala Arg Ala Met Glu Ala Arg Gly Leu Asp Ala Arg
1 5 10 15

<210> 15
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 15
Ala Val Ala Ser Leu Pro Pro Glu Gln Met Phe Glu Leu Met Lys
1 5 10 15

<210> 16
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 16
Ala Met Glu Ala Arg Ala Met Glu Val Arg Gly Met Glu Ala Arg
1 5 10 15

<210> 17
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 17
Gly Tyr Leu Gly Pro Pro His Gln Gly Pro Pro Met His His Val Pro
1 5 10 15

Gly His Glu Ser Arg
20

<210> 18
<211> 24
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 18
Gly Pro Ile Pro Ser Gly Met Gln Gly Pro Ser Pro Ile Asn Met Gly
1 5 10 15

Ala Val Val Pro Gln Gly Ser Arg
20

<210> 19
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 19
Asn Met Leu Leu Gln Asn Pro Gln Leu Ala Tyr Ala Leu Leu Gln Ala
1 5 10 15

Gln Val Val Met Arg
20

<210> 20
<211> 22
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 20
Gly Gly Pro Leu Pro Glu Pro Arg Pro Leu Met Ala Glu Pro Arg Gly
1 5 10 15

Pro Met Leu Asp Gln Arg
20

<210> 21
<211> 28
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:tryptic peptide

<400> 21
Ser Leu Gly Thr Gly Ala Pro Val Ile Glu Ser Pro Tyr Gly Glu Thr
1 5 10 15

Ile Ser Pro Glu Asp Ala Pro Glu Ser Ile Ser Lys
20 25

<210> 22
<211> 500
<212> PRT
<213> Homo sapiens

<400> 22
Met Ala Gly Leu Thr Val Arg Asp Pro Ala Val Asp Arg Ser Leu Arg
1 5 10 15

Ser Val Phe Val Gly Asn Ile Pro Tyr Glu Ala Thr Glu Glu Gln Leu
20 25 30

Lys Asp Ile Phe Ser Glu Val Gly Pro Val Val Ser Phe Arg Leu Val
35 40 45

Tyr Asp Arg Glu Thr Gly Lys Pro Lys Gly Tyr Gly Phe Cys Glu Tyr
50 55 60

Gln Asp Gln Glu Thr Ala Leu Ser Ala Met Arg Asn Leu Asn Gly Arg
65 70 75 80

Glu Phe Ser Gly Arg Ala Leu Arg Val Asp Asn Ala Ala Ser Glu Lys
85 90 95

Asn Lys Glu Glu Leu Lys Ser Leu Gly Thr Gly Ala Pro Val Ile Glu
100 105 110

Ser Pro Tyr Gly Glu Thr Ile Ser Pro Glu Asp Ala Pro Glu Ser Ile
115 120 125

Ser Lys Ala Val Ala Ser Leu Pro Pro Glu Gln Met Phe Glu Leu Met
130 135 140

Lys Gln Met Lys Leu Cys Val Gln Asn Ser Pro Gln Glu Ala Arg Asn
145 150 155 160

Met Leu Leu Gln Asn Pro Gln Leu Ala Tyr Ala Leu Leu Gln Ala Gln
165 170 175

Val Val Met Arg Ile Val Asp Pro Glu Ile Ala Leu Lys Ile Leu His
180 185 190

Arg Gln Thr Asn Ile Pro Thr Leu Ile Ala Gly Asn Pro Gln Pro Val
195 200 205

His Gly Ala Gly Pro Gly Ser Gly Ser Asn Val Ser Met Asn Gln Gln
210 215 220

Asn Pro Gln Ala Pro Gln Ala Gln Ser Leu Gly Gly Met His Val Asn
225 230 235 240

Gly Ala Pro Pro Leu Met Gln Ala Ser Met Gln Gly Gly Val Pro Ala
245 250 255

Pro Gly Gln Met Pro Ala Ala Val Thr Gly Pro Gly Pro Gly Ser Leu
260 265 270

Ala Pro Gly Gly Gly Met Gln Ala Gln Val Gly Met Pro Gly Ser Gly
275 280 285

```
Pro Val Ser Met Glu Arg Gly Gln Val Pro Met Gln Asp Pro Arg Ala
    290                 295                 300

Ala Met Gln Arg Gly Ser Leu Pro Ala Asn Val Pro Thr Pro Arg Gly
305                 310                 315                 320

Leu Leu Gly Asp Ala Pro Asn Asp Pro Arg Gly Gly Thr Leu Leu Ser
                325                 330                 335

Val Thr Gly Glu Val Glu Pro Arg Gly Tyr Leu Gly Pro Pro His Gln
            340                 345                 350

Gly Pro Pro Met His His Val Pro Gly His Glu Ser Arg Gly Pro Pro
        355                 360                 365

Pro His Glu Leu Arg Gly Gly Pro Leu Pro Glu Pro Arg Pro Leu Met
    370                 375                 380

Ala Glu Pro Arg Gly Pro Met Leu Asp Gln Arg Gly Pro Pro Leu Asp
385                 390                 395                 400

Gly Arg Gly Gly Arg Asp Pro Arg Gly Ile Asp Ala Arg Gly Met Glu
                405                 410                 415

Ala Arg Ala Met Glu Ala Arg Gly Leu Asp Ala Arg Gly Leu Glu Ala
            420                 425                 430

Arg Ala Met Glu Ala Arg Ala Met Glu Ala Arg Ala Met Glu Ala Arg
        435                 440                 445

Ala Met Glu Ala Arg Ala Met Glu Val Arg Gly Met Glu Ala Arg Gly
    450                 455                 460

Met Asp Thr Arg Gly Pro Val Pro Gly Pro Arg Gly Pro Ile Pro Ser
465                 470                 475                 480

Gly Met Gln Gly Pro Ser Pro Ile Asn Met Gly Ala Val Val Pro Gln
                485                 490                 495

Gly Ser Arg Gln
            500
```

<210> 23
<211> 577
<212> PRT
<213> Homo sapiens

<400> 23

```
Met Ala Gly Leu Thr Val Arg Asp Pro Ala Val Asp Arg Ser Leu Arg
  1               5                  10                  15

Ser Val Phe Val Gly Asn Ile Pro Tyr Glu Ala Thr Glu Glu Gln Leu
             20                  25                  30

Lys Asp Ile Phe Ser Glu Val Gly Pro Val Val Ser Phe Arg Leu Val
         35                  40                  45

Tyr Asp Arg Glu Thr Gly Lys Pro Lys Gly Tyr Gly Phe Cys Glu Tyr
     50                  55                  60
```

```
Gln Asp Gln Glu Thr Ala Leu Ser Ala Met Arg Asn Leu Asn Gly Arg
 65                  70                  75                  80

Glu Phe Ser Gly Arg Ala Leu Arg Val Asp Asn Ala Ala Ser Glu Lys
                85                  90                  95

Asn Lys Glu Glu Leu Lys Ser Leu Gly Thr Gly Ala Pro Val Ile Glu
            100                 105                 110

Ser Pro Tyr Gly Glu Thr Ile Ser Pro Glu Asp Ala Pro Glu Ser Ile
        115                 120                 125

Ser Lys Ala Val Ala Ser Leu Pro Pro Glu Gln Met Phe Glu Leu Met
    130                 135                 140

Lys Gln Met Lys Leu Cys Val Gln Asn Ser Pro Gln Glu Ala Arg Asn
145                 150                 155                 160

Met Leu Leu Gln Asn Pro Gln Leu Ala Tyr Ala Leu Leu Gln Ala Gln
                165                 170                 175

Val Val Met Arg Ile Val Asp Pro Glu Ile Ala Leu Lys Ile Leu His
            180                 185                 190

Arg Gln Thr Asn Ile Pro Thr Leu Ile Ala Gly Asn Pro Gln Pro Val
        195                 200                 205

His Gly Ala Gly Pro Gly Ser Gly Ser Asn Val Ser Met Asn Gln Gln
    210                 215                 220

Asn Pro Gln Ala Pro Gln Ala Gln Ser Leu Gly Gly Met His Val Asn
225                 230                 235                 240

Gly Ala Pro Pro Leu Met Gln Ala Ser Met Gln Gly Gly Val Pro Ala
                245                 250                 255

Pro Gly Gln Met Pro Ala Ala Val Thr Gly Pro Gly Pro Gly Ser Leu
            260                 265                 270

Ala Pro Gly Gly Gly Met Gln Ala Gln Val Gly Met Pro Gly Ser Gly
        275                 280                 285

Pro Val Ser Met Glu Arg Gly Gln Val Pro Met Gln Asp Pro Arg Ala
    290                 295                 300

Ala Met Gln Arg Gly Ser Leu Pro Ala Asn Val Pro Thr Pro Arg Gly
305                 310                 315                 320

Leu Leu Gly Asp Ala Pro Asn Asp Pro Arg Gly Gly Thr Leu Leu Ser
                325                 330                 335

Val Thr Gly Glu Val Glu Pro Arg Gly Tyr Leu Gly Pro Pro His Gln
            340                 345                 350

Gly Pro Pro Met His His Val Pro Gly His Glu Ser Arg Gly Pro Pro
        355                 360                 365

Pro His Glu Leu Arg Gly Gly Pro Leu Pro Glu Pro Arg Pro Leu Met
    370                 375                 380
```

```
Ala Glu Pro Arg Gly Pro Met Leu Asp Gln Arg Gly Pro Pro Leu Asp
385                 390                 395                 400

Gly Arg Gly Gly Arg Asp Pro Arg Gly Ile Asp Ala Arg Gly Met Glu
                405                 410                 415

Ala Arg Ala Met Glu Ala Arg Gly Leu Asp Ala Arg Gly Leu Glu Ala
            420                 425                 430

Arg Ala Met Glu Ala Arg Ala Met Glu Ala Arg Ala Met Glu Ala Arg
        435                 440                 445

Ala Met Glu Ala Arg Ala Met Glu Val Arg Gly Met Glu Ala Arg Gly
    450                 455                 460

Met Asp Thr Arg Gly Pro Val Pro Gly Pro Arg Gly Pro Ile Pro Ser
465                 470                 475                 480

Gly Met Gln Gly Pro Ser Pro Ile Asn Met Gly Ala Val Val Pro Gln
                485                 490                 495

Gly Ser Arg Gln Val Pro Val Met Gln Gly Thr Gly Met Gln Gly Ala
            500                 505                 510

Ser Ile Gln Gly Gly Ser Gln Pro Gly Gly Phe Ser Pro Gly Gln Asn
        515                 520                 525

Gln Val Thr Pro Gln Asp His Glu Lys Ala Ala Leu Ile Met Gln Val
    530                 535                 540

Leu Gln Leu Thr Ala Asp Gln Ile Ala Met Leu Pro Pro Glu Gln Arg
545                 550                 555                 560

Gln Ser Ile Leu Ile Leu Lys Glu Gln Ile Gln Lys Ser Thr Gly Ala
                565                 570                 575

Pro
```